View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

JOURNAL OF DIFFERENTIAL EQUATIONS 5, 515–530 (1969)

The Cauchy Problem for a Nonlinear First Order Partial Differential Equation

WENDELL H. FLEMING*

Department of Mathematics, Brown University, Providence, Rhode Island 02912

Received August 4, 1967

1. This paper is concerned with the existence and uniqueness of generalized solutions to the Cauchy problem

$$\frac{\partial \phi}{\partial s} + F\left(s, x, \frac{\partial \phi}{\partial x}\right) = 0, \qquad T_0 \leqslant s \leqslant T, \tag{1.1}$$

in one time-like variable s and n space-like variables $x = (x_1, ..., x_n)$, where $\partial\phi/\partial x$ denotes the gradient in the space-like variables, with the terminal data

$$\phi(T, x) = \Phi(x). \tag{1.2}$$

We call ϕ a *generalized solution* of (1.1) if ϕ is bounded, Lipschitz, and satisfies (1.1) almost everywhere in the strip $[T_0, T] \times R^n$. We obtain a generalized solution as the limit when $\epsilon \to 0^+$ of the solution of the corresponding Cauchy problem

$$\frac{\partial \phi^\epsilon}{\partial s} + \frac{\epsilon^2}{2} \Delta_x \phi^\epsilon + F\left(s, x, \frac{\partial \phi^\epsilon}{\partial x}\right) = 0, \qquad T_0 \leqslant s \leqslant T, \tag{1.1$^\epsilon$}$$

$$\phi^\epsilon(T, x) = \Phi^\epsilon(x), \tag{1.2$^\epsilon$}$$

where Φ^ϵ tends to Φ as $\epsilon \to 0^+$ (§'s 4, 5) and Δ_x denotes the Laplace operator in the space-like variables.

When $n = 1$, problem (1.1)–(1.2) is equivalent to

$$\frac{\partial v}{\partial s} + \frac{\partial}{\partial x} F(s, x, v) = 0, \qquad T_0 \leqslant s \leqslant T, \tag{1.3}$$

$$v(T, x) = V(x), \tag{1.4}$$

if we take $v = \partial\phi/\partial x$, $V = \partial\Phi/\partial x$. Problem (1.3)–(1.4) has a long history.

* This work was supported by a grant from the National Science Foundation.

[In the literature usually an initial rather than terminal value Cauchy problem is considered. The substitution $t = T - s$ turns the one problem into the other.] For $n = 1$ our main theorem (Section 5) states that if $v^\epsilon = \partial\phi^\epsilon/\partial x$ is the solution of the Cauchy problem for the corresponding parabolic equation, then $v^\epsilon(s, x)$ tends to $v(s, x)$ at every regular point (Section 3) and hence except on a "small set." Results of this type have been proved by Hopf [*10*], Oleinik [*17*], Donsker [*3*], Varadhan [*18*], and others. However, our assumptions on F are in some respects weaker and when $n = 1$ our main theorem does not seem to be contained in the previous ones. For $n > 1$, (1.3) is replaced by a system of n equations for $v_1, \ldots, v_n$, where $v_i = \partial\phi/\partial x_i$; see Kuznetzov and Šiškin [*15*], Kruzhkov [*13*].

Our method is to write (1.1) as the Hamilton–Jacobi equation of a calculus of variations problem, and (1.1$^\epsilon$) as the Hamilton–Jacobi equation of a corresponding problem in stochastic calculus of variations. By using a probabilistic method, our approach is in principle similar to that of [*3*] and [*18*] but differs from it greatly in detail. The idea of associating a variational problem with (1.1) occurs in Hopf's original paper [*11*], and was exploited systematically by Hopf and Conway [2].

Among the assumptions on $F(s, x, p)$ is a strict concavity condition in p, which corresponds to what Lax [*16*] called genuine nonlinearity of (1.1). Without some such assumption (1.1) is not the Hamilton–Jacobi equation of a variational problem. However, we showed in [*7*] that (1.1) can always be regarded as the Hamilton–Jacobi equation of a certain kind of differential game, provided $F(s, x, p)$ grows no faster than $|p|$ as $|p| \to \infty$. By game theoretic reasoning uniform convergence of ϕ^ϵ to ϕ was proved.

It is known from examples, and from general results of Hopf [*12*], that (1.1)–(1.2) does not have a unique generalized solution. However, there is uniqueness in a subclass of generalized solutions which satisfy an additional condition of Oleinik [*17*] for $n = 1$ and Kruzhkov [*13*] for $n > 1$. In Section 6 we verify that the generalized solution ϕ which we obtain satisfies this condition.

2. *The variational problem.* We assume that $F(s, x, p)$ is of class $C^{(3)}$ and:

(1) $|F_s| + |F_p| + |F_{x_i p_j}| + |F_{s p_j}| \leqslant C(|p|)$,

$$\gamma(|p|)|\lambda|^2 \leqslant -\sum_{i,j=1}^{n} F_{p_i p_j}\lambda_i\lambda_j \leqslant C(|p|)|\lambda|^2$$

for all $\lambda \in R^n$, where $\gamma(v)$, $C(v)$ are positive functions, respectively nonincreasing and nondecreasing in v.

(2) $\lim_{|p|\to\infty} \frac{F(s, x, p)}{|p|} = -\infty.$

Given $r > 0$ there exists k_r such that $|F_p| \leqslant r$ implies $|p| \leqslant k_r$.

(3) $F(s, x, p) \leqslant C$ for some C.

(4) $|F_x| \leqslant c_1(F - p \cdot F_p) + c_2$ for some positive c_1, c_2. Here F_x (or $\partial F/\partial x$) denotes gradient in the variables x, F_p gradient in the variables p, etc.

Example. Let $F(x, p) = -a(x)(1 + |p|^2)^{b/2}$ where $b > 1$, $0 < a_0 \leqslant a(x) \leqslant a_1$, $a'(x)$ is bounded. Then (1)–(4) are satisfied. In place of (4) Oleinik ([*17*], Section 3, page 112) imposed a condition on F_x which is satisfied in this example only if $T - T_0$ is not too large.

Let us assume that the function Φ in (1.2) is of class $C^{(1)}$ and that

$$|\Phi(x)| \leqslant M_0, \qquad |\Phi_x(x)| \leqslant N_0$$

for some M_0, N_0. For some of our results one could assume Φ Lipschitz rather than $C^{(1)}$. However, without some smoothness of Φ as well as F one cannot show the interesting connection of (1.1) with classical calculus of variations.

The variational integrand will be the function $L(s, x, y)$ strictly convex in y which is dual to the function $F(s, x, p)$, which by (1) is strictly concave in p:

$$L(s, x, y) = \max_p [F(s, x, p) - y \cdot p]. \tag{2.1}$$

The function L is $C^{(2)}$ and the dual formula

$$F(s, x, p) = \min_y [L(s, x, y) + y \cdot p] \tag{2.1$'$}$$

holds. The points p and y where the max and min occur are related by the classical Legendre transformation

$$y = F_p, \qquad p = -L_y,$$

which for each (s, x) is globally one-one from R^n onto itself. The matrices L_{yy} and F_{pp} of second partial derivatives are related by $L_{yy} = -(F_{pp})^{-1}$. Moreover,

$$L_x = F_x, \qquad L_s = F_s,$$
$$L_{x_iy} = -F_{x_ip}L_{yy}, \qquad L_{sy} = -F_{sp}L_{yy}.$$

From assumptions (1)–(4) on F we see that

$$\text{(a)} \qquad |L_s| + |L_y| + |L_{x_iy_j}| + |L_{sy_j}| \leqslant K(|y|). \tag{2.2}$$

$$C(k_r)^{-1}|\lambda|^2 \leqslant \sum_{i,j=1}^{n} L_{y_iy_j}\lambda_i\lambda_j \leqslant \gamma(k_r)^{-1}|\lambda|^2 \qquad \text{if} \qquad |y| \leqslant r.$$

(b) $$-\frac{c_2}{c_1} \leqslant L \leqslant C + rk_r \qquad \text{if} \qquad |y| \leqslant r.$$

(c) $$|L_x| \leqslant c_1 L + c_2 .$$

Given (s, x) with $s \leqslant T$ we consider the problem of minimizing

$$J(\xi) = \int_s^T L[t, \xi(t), \dot{\xi}(t)]\, dt + \Phi[\xi(T)] \tag{2.3}$$

among all curves ξ of class $C^{(1)}$ such that $\xi(s) = x$. [We could admit less smooth curves ξ, but this is unnecessary since any minimizing curve must be smooth]. From (2.1′) the Hamilton-Jacobi equation of this minimum problem is (1.1). We consider the minimum value as a function of the initial point (s, x).

THEOREM 1. *Let $\phi(s, x) = \min_\xi J(\xi)$. For any $T_0 < T$, ϕ is bounded and Lipschitz on $[T_0, T] \times R^n$. At each point (s, x) where ϕ is (totally) differentiable, ϕ satisfies* (1.1) *and there is a unique minimizing curve ξ^0 with initial point (s, x).*

Proof. Let us first consider the same minimum problem with the constraint $|\dot{\xi}(t)| \leqslant r$, where $0 < r < \infty$. Since L is convex in y, J is lower semicontinuous under uniform convergence. Therefore by standard reasoning using Ascoli's theorem there is a minimizing curve ξ^0, which is of class $C^{(1)}$ since the matrix L_{yy} is everywhere positive definite. Moreover, if

$$p^0(t) = \int_t^T L_x[\tau, \xi^0(\tau), \dot{\xi}^0(\tau)]\, d\tau + \Phi_x[\xi^0(T)] \tag{2.4}$$

then $L[t, \xi^0(t), y] + y \cdot p^0(t)$ is minimum on $|y| \leqslant r$ if (and only if) $y = \dot{\xi}^0(t)$. This is a special case of Pontryagin's maximum principle for constrained variational problems [*10*]. In this simple problem an easy direct proof can be given.

Let us show that for large enough r the minimum does not actually depend on r. Let

$$\phi^r(s, x) = \min_{|\dot{\xi}| \leqslant r} J(\xi).$$

Now $J(\xi^0) \leqslant J(\xi_x)$ where $\xi_x(t) \equiv x$ for $s \leqslant t \leqslant T$. From this and (2.2), if $T_0 \leqslant s \leqslant T$, then

$$-\left(M_0 + \frac{c_2}{c_1}(T - T_0)\right) \leqslant \phi^r(s, x) \leqslant C(T - T_0) + M_0 .$$

From (2.2c) and (2.4)

$$|p^0(t)| \leqslant \int_t^T (c_1 L + c_2)\, d\tau + N_0 = c_1[\phi^r(t, \xi^0(t)) - \Phi(\xi^0(T))] + c_2(T - t) + N_0 ,$$

$$|p^0(t)| \leqslant (c_1 C + c_2)(T - T_0) + 2c_1 M_0 + N_0 ,$$

which bounds $p^0(t)$ independent of r. Choose r_1 such that

$$|F_p(t, \xi^0(t), p^0(t))| \leqslant r_1 .$$

Now $y^0(t) = F_p(t, \xi^0(t), p^0(t))$ is the point y where $L(t, \xi^0(t), y) + y \cdot p^0(t)$ is minimum on R^n. Hence $y^0(t) = \dot{\xi}^0(t)$ if $r \geqslant r_1$. Thus $|\dot{\xi}^0| \leqslant r_1$ and ξ^0 gives a minimum without the constraint $|\dot{\xi}^0| \leqslant r$. We set $\phi = \phi^r$ for $r \geqslant r_1$.

To show that ϕ is Lipschitz let us show that

(*) $|\phi(s, x') - \phi(s, x)| \leqslant N_1 |x' - x|$

(**) $|\phi(s', x) - \phi(s, x)| \leqslant N_2 |s' - s|$

where N_1 is a bound for $|L_x|$ when $|y| \leqslant r_1$ and $N_2 = B + N_1 r_1$, where B is a bound for $|L|$ when $|y| \leqslant r_1$.

To prove (*) let ξ^0 be as above and consider the translated curve $\xi^0 + x' - x$ with initial point (s, x'). Then

$$|J(\xi^0 + x' - x) - J(\xi^0)| \leqslant N_1 |x' - x|,$$
$$\phi(s, x') \leqslant J(\xi^0 + x' - x), \qquad \phi(s, x) = J(\xi^0),$$
$$\phi(s, x') - \phi(s, x) \leqslant N_1 |x' - x|.$$

By interchanging the roles of x and x' we get the opposite inequality, and hence (*).

To prove (**) we may suppose that $s' > s$. Then

$$\phi(s, x) = \int_s^{s'} L(t, \xi^0, \dot{\xi}^0)\, dt + \phi(s', \xi^0(s')),$$

$$|\phi(s', x) - \phi(s, x)| \leqslant B(s' - s) + N_1 |\xi^0(s') - x|.$$

Since $|\dot{\xi}^0| \leqslant r_1$ we get (**).

If ϕ is differentiable at (s, x), then for any ξ of class $C^{(1)}$ with $\xi(s) = x$,

$$0 \leqslant \phi_s(s, x) + \delta^{-1} \int_s^{s+\delta} [L(s, x, \dot{\xi}(t)) + \phi_x(s, x) \cdot \dot{\xi}(t)]\, dt + 0(1) \quad \text{as} \quad \delta \to 0^+$$

with equality when $\xi = \xi^0$. This implies that if $y = \dot{\xi}(s)$

$$0 \leqslant \phi_s(s, x) + L(s, x, y) + \phi_x(s, x) \cdot y$$

with equality when $y = \dot{\xi}^0(s)$. By (2.1′), equation (1.1) holds at (s, x).

Since L is strictly convex in y, $\dot{\xi}^0(s)$ is the only y at which

$$L(s, x, y) + \phi_x(s, x) \cdot y$$

is minimum. Hence the minimizing curve ξ^0 through a point (s, x) where ϕ is differentiable is unique.

Remarks. The proof that (1.1) holds is nothing but a careful derivation of Bellman's principle of optimality in dynamic programming [*1*]. It is not essential to this derivation that we admit only smooth curves ξ, nor that L is convex in y ([*8*], bottom page 270).

The function ϕ in the theorem clearly satisfies (1.2), and is therefore a generalized solution to the Cauchy problem.

In the course of the proof we have shown that

$$p^0(s) = \phi_x(s, x) \tag{2.5}$$

if ϕ is differentiable at (s, x) and p^0 as in (2.4).

Any minimizing curve ξ^0 is of class $C^{(3)}$ and satisfies Euler's equations together with the transversality condition

$$L_y[T, \xi^0(T), \dot{\xi}^0(T)] = -\phi_x[\xi^0(T)].$$

3. *Regular points.* By Rademacher's theorem, the Lipschitz function ϕ in Theorem 1 is differentiable almost everywhere. However, considerably sharper information about ϕ is available. Following [*15*] let us call (s, x) a *regular point* for the minimum problem (2.3) if there is a unique minimizing curve with initial point (s, x).

THEOREM (*Kuznetzov–Šiškin* [*15*]). *The generalized solution ϕ in Theorem 1 is differentiable at (s, x) if and only if (s, x) is a regular point.*

"Only if" is the last statement of Theorem 1. For completeness let us repeat the proof ([*15*], page 200) of "if." Define $G(s, x, p)$ as $J(\xi)$ for the extremal ξ with initial point (s, x) and initial slope $\dot{\xi}(s) = F_p(s, x, p)$. Let $s, x)$ be regular, ξ^0 the unique minimizing curve through (s, x), and $p^0 = p^0(s)$ as in (2.4). The function G is defined and $C^{(1)}$ in some neighborhood of (s, x, p^0). Let $p^{0'} = p^{0'}(s')$ be defined similarly for a minimizing curve $\xi^{0'}$ through (s', x') with $|\dot{\xi}^0| \leqslant r_1$, $|\dot{\xi}^{0'}| \leqslant r_1$, where r_1 is as in Section 2. As (s', x') tends to (s, x), $\dot{\xi}^{0'}(s')$ tends to $\dot{\xi}^0(s)$ since (s, x) is regular. Hence $p^{0'}$ tends to p^0. Now

$$\begin{aligned} G(s', x', p^{0'}) - G(s, x, p^{0'}) &\leqslant \phi(s', x') - \phi(s, x) \\ &\leqslant G(s', x', p^0) - G(s, x, p^0). \end{aligned}$$

The left side is $G_s^*(s' - s) + G_x^* \cdot (x' - x)$ and the right side

$$G_s^{**}(s' - s) + G_x^{**} \cdot (x' - x),$$

where $*$ means evaluated at $(s^*, x^*, p^{0'})$ and $**$ at (s^{**}, x^{**}, p^0) using the mean value theorem. Therefore ϕ is differentiable at (s, x) and

$$\phi_s(s, x) = G_s(s, x, p^0), \qquad \phi_x(s, x) = G_x(s, x, p^0).$$

Only the initial endpoint of a minimizing extremal ξ^0 can be irregular. For if $(s', \xi^0(s'))$ is irregular for $s < s' < T$, let

$$\xi(t) = \xi^0(t), \qquad s \leqslant t \leqslant s'.$$
$$\xi(t) = \xi^1(t), \qquad s' \leqslant t \leqslant T,$$

where ξ^1 is another minimizing extremal with initial point $(s', \xi^0(s'))$. Then ξ is minimizing and has a corner at $(s', \xi(s'))$, which is impossible since $(L_{y_iy_j})$ is an everywhere positive definite matrix.

By classical methods in calculus of variations we get further information, provided F and Φ are smooth enough. For simplicity let us for the present, and in Theorem 2, assume that F, Φ are of class $C^{(\infty)}$. For each $\alpha \in R^n$ let $\xi(\cdot, \alpha)$ be the extremal curve such that

$$\xi(T, \alpha) = \alpha, \qquad L_y(T, \alpha, \dot{\xi}(T, \alpha)) = -\Phi_x(\alpha), \tag{3.1}$$

the second equation being the transversality condition. The extremal $\xi(\cdot, \alpha)$ is defined and $C^{(\infty)}$ for $s_0(\alpha) < t \leqslant T$, where $-\infty \leqslant s_0(\alpha) < T$. [For the present we ignore T_0 in the previous section.] For each α there is a least number $s_1(\alpha) \geqslant s_0(\alpha)$ such that $\xi(\cdot, \alpha)$ minimizes $J(\xi)$ among all ξ with $\xi(s) = \xi(s, \alpha)$ provided $s \geqslant s_1(\alpha)$. If $s_1(\alpha) = s_0(\alpha) = -\infty$ this is true for any $s \leqslant T$. Every minimizing curve coincides with a final portion of $\xi(\cdot, \alpha)$ for some α.

Let us call (s, x) a *conjugate point* for the extremal $\xi(\cdot, \alpha)$ if

$$x = \xi(s, \alpha), \qquad \frac{\partial \xi}{\partial \alpha}(s, \alpha) = 0,$$

where $\partial\xi/\partial\alpha$ is the Jacobian. By a classical result, only the initial endpoint of a minimizing curve can be a conjugate point ([*10*], Chapter 3). Let

$$E_1 = \left\{(s_1(\alpha), x) : x = \xi(s_1(\alpha), \alpha), \frac{\partial \xi}{\partial \alpha}(s_1(\alpha), \alpha) = 0\right\},$$
$$E_2 = \{\text{all irregular points not in } E_1\},$$
$$E_{is} = \{x : (s, x) \in E_i\}, \qquad i = 1, 2.$$

The function ϕ in Section 2 is defined in the half-space $H = (-\infty, T) \times R^n$.

THEOREM 2. (a) *$E_1 \cup E_2$ is a closed set.*

(b) *ϕ is of class $C^{(\infty)}$ on $H - (E_1 \cup E_2)$.*

(c) *For each s and $\theta > 0$, $m^{n-1+\theta}(E_{1s}) = 0$, where m^q is Hausdorff q-dimensional measure.*

(d) *For each s and $x \in E_{2s}$ there is a neighborhood U of x such that $E_{2s} \cap U$ is contained in a finite union of $(n - 1)$-manifolds of class $C^{(\infty)}$.*

Proof. By the classical method of characteristics for the Hamilton–Jacobi equation, $E_1 \cup E_2$ can have no limit point (s, x) with $s = T$. If $(s, x) \notin E_1 \cup E_2$, $s < T$, then there exists α_0 such that

$$x = \xi(s, \alpha_0), \qquad \frac{\partial \xi}{\partial \alpha}(t, \alpha_0) \neq 0 \qquad \text{for} \quad s \leqslant t \leqslant T,$$

and $\xi(\cdot, \alpha_0)$ is the unique minimizing curve with initial point (s, x). For α in some neighborhood of α_0, the curves $\xi(\cdot, \alpha)$ form a field ([*10*], Chapter 3) covering simply a neighborhood of this minimizing curve. For (s', x') near (s, x) any minimizing curve $\xi^{0\prime}$ with initial point (s', x') must have $\dot{\xi}^{0\prime}(s')$ near $\dot{\xi}^0(s)$. Hence $\xi^{0\prime}$ must coincide with some $\xi(\cdot, \alpha)$ of the field for $s' \leqslant t \leqslant T$. Hence $(s', x') \notin E_1 \cup E_2$, which shows that $E_1 \cup E_2$ is closed. Moreover, the method of characteristics shows that ϕ is $C^{(\infty)}$ in a neighborhood of (s, x). This proves (a) and (b).

Part (c) follows from known results (see [*5*] and references cited there), applied to the mapping $\alpha \to \xi(s, \alpha)$. To prove (d), let

$$J(s, \alpha) = \int_s^T L[t, \xi(t, \alpha), \dot{\xi}(t, \alpha)]\, dt + \Phi(\alpha). \tag{3.2}$$

Consider the mapping $\Gamma : R^n \times R^n \to R^{n+1}$ such that (for given s)

$$\Gamma(\alpha, \alpha') = (\xi(s, \alpha) - \xi(s, \alpha'), J(s, \alpha) - J(s, \alpha')).$$

By differentiating (3.2) with respect to α_i and using (3.1) we find that

$$\frac{\partial J}{\partial \alpha_i} = -L_y(s, \xi(s, \alpha), \dot{\xi}(s, \alpha)) \frac{\partial \xi}{\partial \alpha_i}. \tag{3.3}$$

Let $(s, x) \in E_2$. Then for distinct α_0, α_0',

$$\xi(s, \alpha_0) = \xi(s, \alpha_0') = x$$
$$J(s, \alpha_0) = J(s, \alpha_0') = \phi(s, x),$$

and $\partial\xi/\partial\alpha \neq 0$ at (s, α_0), (s, α_0'). If the differential $d\Gamma$ has rank less than $n + 1$ at (α_0, α_0'), then using (3.3) we find that

$$L_y(s, x, \dot{\xi}(s, \alpha_0)) = L_y(s, x, \dot{\xi}(s, \alpha_0')),$$

which is impossible since the Legendre transformation is one–one. Thus $d\Gamma$ has maximum rank $n + 1$ at (α_0, α_0'), which implies that near (α_0, α_0') the set of (α, α') where $\Gamma(\alpha, \alpha') = 0$ is an $(n - 1)$-manifold of class $C^{(\infty)}$. For each $(s, x) \in E_2$ there are only finitely many such pairs (α_0, α_0'), since all minimizing extremals have uniformly bounded slopes on $[s, T]$. This proves (d).

Note. If F is of class $C^{(k+2)}$ and Φ of class $C^{(k+1)}$, then ξ is $C^{(k)}$ and [5] $m^{n-1+1/k}(E_{1s}) = 0$. If F and Φ are real analytic, then any bounded subset of E_{1s} has finite m^{n-1} measure.

The following lemma is useful in proving the main theorem in Section 5. Let us use the notation

$$\| \xi \| = \max_{s \leqslant t \leqslant T} | \xi(t)|.$$

LEMMA. *Let (s, x) be regular, and ξ^0 the unique minimizing curve with initial point (s, x). Then given $r > 0$ and $\theta > 0$ there exists $\delta > 0$ with the following property: if $\| \dot{\xi} \| \leqslant r$ and $J(\xi) < J(\xi^0) + \delta$, then $\| \xi - \xi^0 \| < \theta$ and $\int_s^T | \dot{\xi} - \dot{\xi}^0 |^2 \, dt < \theta$.*

Proof. If for $k = 1, 2,...$

$$\xi_k(s) = x, \qquad \| \dot{\xi}_k \| \leqslant r,$$
$$\lim_{k \to \infty} J(\xi_k) = J(\xi^0),$$

then $\| \xi_k - \xi^0 \| \to 0$ as $k \to \infty$. Otherwise there would exist a subsequence for which ξ_k tends to a different minimizing extremal, contrary to the assumption that (s, x) is regular. This proves all but the last assertion.

Let $| L_x | \leqslant N$ when $| y | \leqslant r$. Then

$$| L(t, \xi, \dot{\xi}) - L(t, \xi^0, \dot{\xi})| \leqslant N \| \xi - \xi^0 \|,$$
$$L(t, \xi^0, \dot{\xi}) - L(t, \xi^0, \dot{\xi}^0) \geqslant L_y(t, \xi^0, \dot{\xi}^0) \cdot (\dot{\xi} - \dot{\xi}^0) + c \| \dot{\xi} - \dot{\xi}^0 \|^2,$$

where $c = \frac{1}{2} C(k_r)^{-1}$ according to (2.2a). Moreover, since ξ^0 is an extremal

$$\int_s^T L_y \cdot (\dot{\xi} - \dot{\xi}^0) \, dt = L_y \cdot (\xi - \xi^0) \Big|_s^T - \int_s^T L_x \cdot (\xi - \xi^0) \, dt$$

from which

$$\left| \int_s^T L_y \cdot (\dot{\xi} - \dot{\xi}^0) \, dt \right| \leqslant K_1 \| \xi - \xi^0 \|$$

for some K_1. From these inequalities we get

$$J(\xi) - J(\xi^0) > c \int_s^T | \dot{\xi} - \dot{\xi}^0 |^2 \, dt - K_2 \| \xi - \xi^0 \|$$

for some positive K_2, from which the lemma follows.

4. *The Cauchy problem* (1.1$^\epsilon$)–(1.2$^\epsilon$). Let us choose Φ^ϵ for $\epsilon > 0$ such that Φ^ϵ is of class $C^{(\infty)}$, $| \Phi^\epsilon | \leqslant M_0$, $| \Phi_x^\epsilon | \leqslant N_0$, the higher order partial

derivatives are bounded by constants possibly depending on ϵ, and Φ^ϵ, Φ_x^ϵ tend uniformly to Φ, Φ_x as $\epsilon \to 0^+$. For instance, we can take $\Phi^\epsilon = \Phi * h^\epsilon$, where h^ϵ is a $C^{(\infty)}$ approximate identity.

The Cauchy problem (1.1^ϵ)–(1.2^ϵ) has a solution ϕ^ϵ which can be represented in terms of a stochastic variational problem corresponding to the one in Section 2, as follows. We consider n-dimensional stochastic processes ξ on the time interval $[s, T]$ of the form

$$\xi(t) = \eta(t) + \epsilon[w(t) - w(s)], \tag{4.1}$$

where η is a process with sample paths of class $C^{(1)}$, $\eta(s) = x$, and w is an n-dimensional Brownian motion. Actually, we should write $\xi(t, \omega)$, $\eta(t, \omega)$, $w(t, \omega)$, where ω belongs to some probability space Ω; following custom we do not exhibit the dependence of processes on ω. We require that ξ be nonanticipative; this means that the random variables $\xi(r)$ for $r \leqslant t$ are independent of Brownian increments for times $\geqslant t$. Instead of (2.3) we now seek to minimize

$$J^\epsilon(\xi) = E\left\{\int_s^T L[t, \xi(t), \dot\eta(t)]\, dt + \Phi^\epsilon[\xi(T)]\right\} \tag{4.2}$$

where $E\{\ \}$ denotes expected value. In fact, we shall restrict ξ to be an n-dimensional Markov process on $[s, T]$, starting at (s, x) such that

$$\dot\eta(t) = Y[t, \xi(t)]. \tag{4.3}$$

We require that $Y \in \mathscr{Y}$, where $\mathscr{Y}$ is the class of R^n-valued functions on the strip $[T_0, T] \times R^n$ which are bounded and satisfy for some positive $M(= M_Y)$ and $\alpha(= \alpha_Y)$:

$$| Y(s, x') - Y(s, x)| \leqslant M \,| x' - x |$$
$$| Y(s', x) - Y(s, x)| \leqslant M \,| s' - s |^\alpha$$

for all $s, s' \in [T_0, T]$ and $x, x' \in R^n$. The minimum value of $J^\epsilon(\xi)$ is the same in the wider class of nonanticipative ξ described above ([8], Theorem 3.1), but we do not use this fact here.

Formulas (4.1), (4.3) are equivalent to the vector stochastic differential equation

$$d\xi(t) = Y(t, \xi(t))\, dt + \epsilon dw(t), \qquad s \leqslant t \leqslant T \tag{4.4}$$

with initial data $\xi(s) = x$. We seek to minimize $J^\epsilon(\xi)$ by suitable choice of the drift coefficient Y in (4.4).

Let $\mathscr{F}$ denote the class of real-valued functions ψ such that $D\psi$ is bounded and Hölder continuous in the strip $[T_0, T] \times R^n$, where $D\psi$ denotes ψ or any of its partial derivatives ψ_s, ψ_{x_i}, $\psi_{x_i x_j}$. Let

$$\begin{aligned} \phi^\epsilon(s, x) &= \min_{Y \in \mathscr{Y}} J^\epsilon(\xi) \\ Y^\epsilon(s, x) &= F_p(s, x, \phi_x^\epsilon(s, x)), \end{aligned} \tag{4.5}$$

and ξ^ϵ, η^ϵ given by (4.3), (4.4) with $Y = Y^\epsilon$. We will show that the minimum is attained in the course of proving:

THEOREM. (a) *ϕ^ϵ is the unique solution in $\mathscr{F}$ of the Cauchy problem* (1.1^ϵ)–(1.2^ϵ).

(b) *Y^ϵ minimizes* (4.2), *simultaneously for all* (s, x) *in the strip* $[T_0, T] \times R^n$.

(c) $\phi_x^\epsilon(s, x) = E\left\{\int_s^T L_x[t, \xi^\epsilon(t), \eta^\epsilon(t)]\, dt + \Phi_x^\epsilon[\xi^\epsilon(T)]\right\}$.

Proof. The proof of (a), (b) is similar to reasoning in ([*6*]; [*8*], Section 7). We first impose the constraint $|Y(s, x)| \leqslant r$ and let

$$F^r(s, x, p) = \min_{|y| \leqslant r} [L(s, x, y) + y \cdot p].$$

Then F^r is bounded and Lipschitz on $[T_0, T] \times R^n \times (|p| \leqslant C)$ for any C, and for suitable K_r

$$|F_p^r| \leqslant K_r, \qquad |F_x^r| \leqslant K_r(1 + |p|).$$

Therefore, the Cauchy problem

$$\begin{aligned} \phi_s^{\epsilon r} + \frac{\epsilon^2}{2} \Delta_x \phi^{\epsilon r} + F^r(s, x, \phi_x^{\epsilon r}) = 0, \qquad T_0 \leqslant s \leqslant T, \\ \phi^{\epsilon r}(T, x) = \Phi^\epsilon(x) \end{aligned}$$

has a unique solution in $\mathscr{F}$ ([*14*], Theorem 14). Let us show that $\phi^{\epsilon r}$ and $\phi_x^{\epsilon r}$ satisfy the same bounds (independent of ϵ and r) as in the proof of Theorem 1. By the reasoning there

$$-\left(M_0 + \frac{c_2}{c_1}(T - T_0)\right) \leqslant \phi^{\epsilon r}(s, x) \leqslant C(T - T_0) + M_0 .$$

Let $Y^*(s, x)$ be the unique point in $|y| \leqslant r$ at which $L(s, x, y) + y \cdot \phi_x^{\epsilon r}(s, x)$ is minimum. From (2.2a) and the fact that $\phi^{\epsilon r} \in \mathscr{F}$, $L_y + \phi_x^{\epsilon r}$ satisfies a uniform Lipschitz condition in x and uniform Hölder condition in s for $|y| \leqslant r$.

Then $Y^* \in \mathscr{Y}$ ([8], Lemma 2.1). For any $Y \in \mathscr{Y}$, $J^\epsilon(\xi) = \psi(s, x)$ where ψ is the unique solution in $\mathscr{F}$ of the linear parabolic equation

$$\psi_s + \frac{\epsilon^2}{2} \Delta_x \psi + L(s, x, Y(s, x)) + Y(s, x) \cdot \psi_x = 0, \qquad T_0 \leqslant s \leqslant T \tag{4.6}$$

with $\psi(T, x) = \Phi^\epsilon(x)$. From the maximum principle for parabolic equations it then follows at once that $J^\epsilon(\xi) \geqslant \phi^{\epsilon r}(s, x)$ for all $Y \in \mathscr{Y}$ such that $| Y | \leqslant r$, with equality when $Y = Y^*$.

To estimate $\phi_x^{\epsilon r}$, consider x, x' and let

$$Y(s, \cdot) = Y^*(s, \cdot + x - x').$$

If ξ^* and ξ' are the solutions of (4.4) corresponding to Y^* and Y with respective initial data

$$\xi^*(s) = x, \qquad \xi'(s) = x',$$

then $\dot\eta' = \dot\eta^*$ and

$$\xi'(t) = \xi^*(t) + x' - x, \qquad s \leqslant t \leqslant T,$$

$$\phi^{\epsilon r}(s, x) = E\left\{\int_s^T L[t, \xi^*, \dot\eta^*]\, dt + \Phi^\epsilon[\xi^*(T)]\right\},$$

$$\phi^{\epsilon r}(s, x') \leqslant E\left\{\int_s^T L[t, \xi', \dot\eta']\, dt + \Phi^\epsilon[\xi'(T)]\right\}.$$

If we subtract and let $x' \to x$, then ([8], page 274)

$$\limsup_{x' \to x} \frac{\phi^{\epsilon r}(s, x') - \phi^{\epsilon r}(s, x)}{| x' - x |} \leqslant E \int_s^T | L_x(t, \xi^*, \dot\eta^*)|\, dt + N_0,$$

$$| \phi_x^{\epsilon r}(s, x)| \leqslant c_1[\phi^{\epsilon r}(s, x) - E\Phi^\epsilon(\xi^*(T))] + c_2(T - s) + N_0,$$

$$| \phi_x^{\epsilon r} | \leqslant (c_1 C + c_2)(T - T_0) + 2c_1 M_0 + N_0$$

independent of r and ϵ. Just as in Section 2 there exists r_1 such that for, $r \geqslant r_1$, $\phi^{\epsilon r} = \phi^\epsilon$ is a solution of (1.1^ϵ). Moreover, $Y^* = Y^\epsilon$ for $r \geqslant r_1$.

To prove (c), we note that ϕ^ϵ is of class $C^{(3)}$ since F is $C^{(3)}$ and Φ^ϵ is $C^{(\infty)}$. Take $\partial/\partial x_i$ in (1.1^ϵ). Then

$$(\phi_{x_i}^\epsilon)_s + \frac{\epsilon^2}{2} \Delta_x \phi_{x_i}^\epsilon + F_p \cdot (\phi_{x_i}^\epsilon)_x = -F_{x_i}.$$

The left side is the backward operator applied to $\phi_{x_i}^\epsilon$ of the Markov process ξ^ϵ

determined through (4.4) by Y^ϵ and the initial data (s, x) [in the notation above, $\xi^* = \xi^\epsilon$]. Hence

$$\phi^\epsilon_{x_i}(s, x) = E\left\{\int_s^T F_{x_i}(t, \xi^\epsilon, p^\epsilon)\, dt + \Phi^\epsilon_{x_i}[\xi^\epsilon(T)]\right\}$$

where $p^\epsilon(t) = \phi_x{}^\epsilon(t, \xi^\epsilon(t))$. But $p^\epsilon(t)$ and $\dot{\eta}^\epsilon(t)$ are dually related through the Legendre transformation. Hence

$$L_x[t, \xi^\epsilon(t), \dot{\eta}^\epsilon(t)] = F_x[t, \xi^\epsilon(t), p^\epsilon(t)]$$

for $s \leqslant t \leqslant T$, which proves (c).

Note. Y^ϵ is the unique function in $\mathscr{Y}$ minimizing (4.2) simultaneously for all (s, x) in the strip. For if $Y \in \mathscr{Y}$ is any such function, then $\psi(s, x) = \phi^\epsilon(s, x)$ for all (s, x). This implies that $L(s, x, y) + y \cdot \phi_x{}^\epsilon(s, x)$ is minimum at both $Y(s, x)$ and $Y^\epsilon(s, x)$. Thus $Y = Y^\epsilon$.

5. *Main Theorem.* Using the representation of $\phi(s, x)$ as the minimum value of $J(\xi)$ and of $\phi^\epsilon(s, x)$ as that of $J^\epsilon(\xi)$, it is now rather easy to obtain the following result.

THEOREM 4. *As* $\epsilon \to 0^+$:

(a) $\phi^\epsilon(s, x)$ *tends uniformly to* $\phi(s, x)$.

(b) $\phi_x{}^\epsilon(s, x)$ *tends to* $\phi_x(s, x)$ *at every regular point* (s, x).

Proof. We use the following observations. Given (s, x), let ξ^0 minimize $J(\xi)$, and ξ^ϵ the Markov process at the end of Section 4 minimizing $J^\epsilon(\xi)$. Consider the process ξ given by

$$\xi(t) = \xi^0(t) + \epsilon[w(t) - w(s)],$$

where in (4.4) we take $Y(t) = \dot{\xi}^0(t)$. Then

$$\phi^\epsilon(s, x) = J^\epsilon(\xi^\epsilon) \leqslant J^\epsilon(\xi). \tag{5.1}$$

On the other hand,

$$\xi^\epsilon(t) = \eta^\epsilon(t) + \epsilon[w(t) - w(s)],$$

where with probability 1, η^ϵ is a $C^{(1)}$ curve through (s, x) and hence

$$\phi(s, x) = J(\xi^0) \leqslant J(\eta^\epsilon). \tag{5.2}$$

From (5.1) and (5.2),

$$\begin{aligned} J^\epsilon(\xi^\epsilon) - EJ(\eta^\epsilon) &\leqslant \phi^\epsilon(s, x) - \phi(s, x) \\ &\leqslant J^\epsilon(\xi) - J(\xi^0). \end{aligned} \tag{5.3}$$

Now

$$J^\epsilon(\xi^\epsilon) - EJ(\eta^\epsilon) = E\left\{\int_s^T [L(t, \xi^\epsilon, \dot\eta^\epsilon) - L(t, \eta^\epsilon, \dot\eta^\epsilon)]\,dt\right.$$

$$\left. + \Phi^\epsilon[\xi^\epsilon(T)] - \Phi^\epsilon[\eta^\epsilon(T)]\right\},$$

$$|J^\epsilon(\xi^\epsilon) - EJ(\eta^\epsilon)| \leqslant [N_1(T-s) + N_0]\,\epsilon E\| w \|,$$

where $\| w \| = \max\{| w(t)| : s \leqslant t \leqslant T\}$ and $| L_x | \leqslant N_1$ whenever $| y | \leqslant r_1$. Since $E\| w \| < \infty$, $J^\epsilon(\xi^\epsilon) - EJ(\eta^\epsilon)$ tends uniformly to 0 as $\epsilon \to 0^+$. By similar reasoning the right side of (5.3) tends uniformly to 0 as $\epsilon \to 0^+$. This proves (a).

To prove (b), let (s, x) be regular. By (2.4), (2.5) and Theorem (3c),

$$\phi_x^\epsilon(s, x) - \phi_x(s, x) = E\left\{\int_s^T [L_x(t, \xi^\epsilon, \dot\eta^\epsilon) - L_x(t, \xi^0, \dot\xi^0)]\,dt\right.$$

$$\left. + \Phi_x^\epsilon[\xi^\epsilon(T)] - \Phi_x[\xi^0(T)]\right\}. \tag{5.4}$$

Since $J(\eta^\epsilon) \geqslant J(\xi^0)$ and $EJ(\eta^\epsilon)$ tends to $J(\xi^0)$, $J(\eta^\epsilon)$ tends to $J(\xi^0)$ in probability as $\epsilon \to 0^+$. By the lemma in Section 3,

$$\| \eta^\epsilon - \xi^0 \| + \int_s^T | \dot\eta^\epsilon - \dot\xi^0 |^2\,dt$$

tends to 0 in probability as $\epsilon \to 0^+$. Moreover,

$$\| \xi^\epsilon - \eta^\epsilon \| = \epsilon \| w \|$$

tends to 0 with probability 1. Therefore

$$A_\epsilon(t) = L_x(t, \xi^\epsilon, \dot\eta^\epsilon) - L_x(t, \xi^0, \dot\xi^0)$$

tends to 0 in probability for each t; and $| A_\epsilon(t)| \leqslant 2N_1$ since $| \dot\eta^\epsilon |$ and $| \dot\xi^0 |$ are bounded by r_1 (Sections 2, 4). Since Φ_x^ϵ tends to Φ_x uniformly, $\Phi_x^\epsilon[\xi^\epsilon(T)]$ tends to $\Phi_x[\xi^0(T)]$ in probability as $\epsilon \to 0^+$. From (5.4) part (b) then follows.

6. *Uniqueness.* Let us say that a generalized solution ϕ of (1.1)–(1.2) has property $\mathscr{L}$ if there is a constant K such that

$$\frac{\Delta_2\phi}{| l |^2} = \frac{\phi(s, x + l) - 2\phi(s, x) + \phi(s, x - l)}{| l |^2} \leqslant \frac{K}{T - s}$$

for $T_0 \leqslant s \leqslant T$ and every $x, l \in R^n$. Besides properties (1)–(4) of F in Section 2 let us also assume that, for $i, j = 1,..., n$,

$$|F_{x_i x_j}| \leqslant C(|p|). \tag{5.5}$$

Then one has the following uniqueness theorem.

THEOREM (Kruzhkov [13]). *The Cauchy problem* (1.1)–(1.2) *has at most one generalized solution with property* $\mathscr{L}$.

To show that there is such a generalized solution we prove:

THEOREM 5. *The function* ϕ *in Theorem* 1 *has property* $\mathscr{L}$.

Proof. Following Kruzhkov [13], who considered $F = F(p)$, it suffices to show that $W = (T - s)\,\phi^\epsilon_{x_i x_i}$ is bounded above by some positive constant K. Now

$$0 = W_s + \frac{\epsilon^2}{2}\Delta_x W + F_p \cdot W_x + \frac{W}{T-s} + \Psi,$$

$$\Psi = (T-s)\left[F_{x_i x_i} + 2\sum_j F_{x_i p_j}\frac{\partial \phi^\epsilon_{x_i}}{\partial x_j} + \sum_{j,k} F_{p_j p_k}\frac{\partial \phi^\epsilon_{x_i}}{\partial x_j}\frac{\partial \phi^\epsilon_{x_i}}{\partial x_k}\right].$$

For suitable $a, b > 0$,

$$\Psi \leqslant (T-s)\left[a - b\sum_j \left(\frac{\partial \phi^\epsilon_{x_i}}{\partial x_j}\right)^2\right] \leqslant (T-s)[a - b(\phi^\epsilon_{x_1 x_1})^2],$$

$$0 \leqslant W_s + \frac{\epsilon^2}{2}\Delta_x W + F_p \cdot W_x + (T-s)\,a + \frac{W - bW^2}{T-s}.$$

Choose K such that, for $T_0 \leqslant s \leqslant T$,

$$(T-s)^2 a + W - bW^2 < 0 \qquad \text{if} \quad W > K.$$

Then $W \leqslant K$. This follows from the maximum principle for parabolic equations, or from the following probabilistic argument. Suppose that $W(s, x) > K$. Now

$$\mathfrak{a}W = W_s + \frac{\epsilon^2}{2}\Delta_x W + F_p \cdot W_x,$$

where $\mathfrak{a}$ is the backward operator of the (strong) Markov process ξ^ϵ with drift coefficient $Y^\epsilon = F_p$, starting at (s, x). Let τ be the (random) time when first $W(t, \xi^\epsilon(t)) = K$. Then ([4], Chapter 13)

$$0 > EW\Big|_s^\tau = E\int_s^\tau \mathfrak{a} W\,dt \geqslant 0,$$

a contradiction.

References

1. Bellman, R., "Dynamic Programming." Princeton, 1957.
2. Conway, E. D. and Hopf, E., Hamilton's theory and generalized solutions of the Hamilton-Jacobi equation. *J. Math. Mech.* **13** (1964), 939–986.
3. Donsker, M., An application of function space integrals to a nonlinear equation. To appear.
4. Dynkin, E. B., "Markov Processes." Springer, Berlin, 1965. [English translation].
5. Federer, H., Two theorems in geometric measure theory. *Bull. Am. Math. Soc.* **72** (1966), 719.
6. Fleming, W. H., Some Markovian optimization problems. *J. Math. Mech.* **12**, No. 1 (1963), 131–140.
7. Fleming, W. H., The Cauchy problem for degenerate parabolic equations. *J. Math. Mech.* **13**, No. 1 (1964), 987–1008.
8. Fleming, W. H., Duality and a priori estimates in Markovian optimization problems. *J. Math. Anal. Appl.* **16**, No. 2 (1966), 254–279.
9. Friedman, A., "Partial Differential Equations of Parabolic Type." Prentice Hall, Englewood Cliffs, 1964.
10. Hestenes, M. R., "Calculus of Variations and Optimal Control Theory." Wiley, New York, 1966.
11. Hopf, E., The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure Appl. Math.* **3** (1950), 201–230.
12. Hopf, E., Generalized solutions of nonlinear equations of first order. *J. Math. Mech.* **14** (1965), 951–972.
13. Kruzhkov, S. N., The Cauchy problem in the large for certain quasilinear systems of first-order with several variables. *Dokl. Akad. Nauk.* **155** (1964), 743–746; *Soviet Math.* **5**, 493–496.
14. Oleĭnik, O. A. and Kruzhkov, S. N., Quasilinear parabolic equations of second order with many independent variables. *Uspekhi Mat. Nauk.* **16**, No. 5 (1961), 115–155.
15. Kuznetzov, N. N. and Šiškin, A. A., On a many dimensional problem in the theory of quasilinear equations. *Z. Vyčisl. Mat. i Mat. Fiz.* **4** (1964), No. 4, suppl. pp. 192–205.
16. Lax, P. D., Hyperbolic systems of conservation laws II. *Comm. Pure Appl. Math.* **10** (1957), 537–566.
17. Oleĭnik, O. A., Discontinuous solutions of non-linear differential equations. *Uspekhi Mat. Nauk.* (NS) **12** (1957), No. 3, pp. 3–73; English transl. *Am. Math. Soc. Transl.*, Ser. 2, **26**, 95–172.
18. Varadhan, S. R. S., Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** (1966), 261–286.